幻觉主义意识理论的解释进路

冯璇

（格拉斯哥大学 英国 苏格兰 格拉斯哥 G12 8QQ）

摘 要：本文考察当代幻觉主义如何由对特殊现象属性的否定，发展为解释现象意识为何显得存在的积极研究进路。文章从现象意识与通达意识的区分及意识困难问题出发，讨论丹尼特对内省证据的批评和弗兰克什对“幻觉问题”的界定，并比较注意图式理论、推理理论与卡默勒元笛卡尔主义分别对自我模型、现象属性归因和第一人称显然性的解释。本文认为，这些方案使幻觉主义获得了较强的解释力，并为意识研究提供了可继续发展的替代框架。不过，现象判断所得到自然主义解释，并不足够支撑相应现象属性不存在的证明；认识论祛魅的效力及其本体论范围，仍需要进一步论证。

关键词：幻觉主义；现象意识；幻觉问题；第一人称显然性；认识论祛魅

DOI：10.64549/jtii.v1i5.105

作者简介：冯璇（2000-），女，格拉斯哥大学 25 届心灵哲学与心理学硕士，研究方向为文艺理论、心灵哲学、归纳推理。

The Explanatory Programme of Illusionism about Consciousness

Xuan Feng

(University of Glasgow, Glasgow, Scotland, United Kingdom, G12 8QQ)

Abstract: This article examines how contemporary illusionism has developed from a denial of special phenomenal properties into a positive explanatory approach to why phenomenal consciousness seems to exist. Beginning with the distinction between phenomenal and access consciousness and the hard problem, it discusses Dennett's criticism of introspective evidence and Frankish's formulation of the illusion problem. It then compares attention schema theory, inference theory, and Kammerer's Meta-Cartesianism in terms of their accounts of self-modeling, phenomenal-property attribution, and the apparent obviousness of first-person experience. These approaches give illusionism considerable explanatory power and make it a promising alternative research programme. Yet a naturalistic explanation of phenomenal judgments does not by itself establish that phenomenal properties do not exist. Whether epistemic debunking can support that stronger ontological conclusion remains an open question.

Keywords: Illusionism; Phenomenal Consciousness; Illusion Problem; First-Person Obviousness; Epistemic Debunking

引言

我们常凭行为判断其他人是否有意识：一个人抱怨疼痛，我们通常相信他正在受苦；动物虽然无法用语言报告，却会回避伤害；人工智能甚至能够细致描述自己的“感受”。这些表现都可以成为判断依据，但仅从信息处理和语言报告出发，我们仍难确定一个系统是否具有主观经验。纳格尔（Thomas Nagel）借蝙蝠的例子指出，即使我们掌握了回声定位的全部生理机制，也未必因而知道以蝙蝠的方式感知世界是什么样的。[1]

现代意识哲学把这一困难集中到现象意识（phenomenal consciousness）上，即经验对主体而言“像什么”（what it is like）的方面。查尔默斯（David J. Chalmers）的困难问题进一步要求解释：为什么物理和认知过程还会伴随这样的经验？这一提问似乎有一个无法绕开的起点——我们可以怀疑其他主体是否有经验，却很难怀疑自己的疼痛、颜色经验正在以某种方式显现。第一人称经验于是被赋予特殊的认识论地位。

幻觉主义（illusionism）对此提出了一种颇为激进的挑战。它不否认人会疼痛、辨认颜色、形成判断或报告自身状态；它质疑的是，我们能否从这些事实，尤其从经验“显得具有某种性质”，直接推出传统意义上的现象属性（phenomenal properties）真实存在。主体可能准确地知道自己处于某种心理状态，却误解了这种状态的性质。幻觉主义因而把内省的可靠性也纳入考察。

弗兰克什（Keith Frankish）提出的“幻觉问题”（illusion problem），使这种本体论争论获得了积极的解释方向：如果传统现象属性并未实例化，为什么我们仍如此稳定地相信它们存在？近期理论又尝试解释这一确信为何显得直接、自明。本文沿着解释对象的变化梳理有关争论，考察不同方案能够说明什么，以及从现象判断的自然主义解释走向现象属性的否定，为何还具有一些距离。

一、现象意识如何成为独立的解释对象

（一）现象意识、通达意识与实验指标

布洛克（Ned Block）区分现象意识与通达意识（access consciousness）：前者涉及经验的主观性质，后者涉及某一心理内容能否被用于推理、报告和行动控制。[2] 二者通常相伴，却没有理由仅凭这种相伴就将其等同。盲视（blindsight）提供了一个有启发性的例子：部分视觉皮层受损患者报告自己在特定视野中看不见刺激，却能以高于随机水平的正确率猜测刺激位置或方向。[3]

盲视不能简单证明患者“有通达意识而无现象意识”，因为有关视觉信息也未必获得正常的认知通达。它说明信息加工、认知利用与主观经验应分别研究。在意识实验中，这一区分同样重要：语言报告常常包含记忆、判断和行为输出，无报告范式虽然试图排除报告环节，却又依赖眼动等代理指标推断经验是否发生。[4] 施利希特与菲利普斯所讨论的方法论难题，正在于如何确保这些指标测量到的是现象意识，而非与之相伴的其他过程。[5,6]

（二）困难问题与现象属性

查尔默斯将注意、记忆、刺激辨别和行为控制等问题称为意识的“容易问题”（easy problems），因为原则上可以通过功能分析与机制研究解释它们；困难问题（hard problem）则追问，这些机制为何还伴随主观经验。[7,8] “容易”并不意味着实验工作简单，区别在于两类问题要求何种解释。

哲学僵尸思想实验凸显了这一差异。设想一个在物理结构、行为和认知功能上与人完全相同，却没有现象经验的存在者。如果这种情形不仅可设想，而且在形而上学上可能，物理事实便不能必然蕴含现象事实。[8] 论证是否成立，取决于可设想性向形而上学可能性的过渡能否得到辩护。

不过，查尔默斯并未因此割裂意识与功能组织。他在渐逝感受质和跳舞感受质的论证中指出：若精细功能组织保持不变而经验逐渐消失或剧烈变化，主体却始终无法在判断和行为上察觉，便会产生难以接受的后果；他据此支持组织不变性。[9] 在这一立场下，功能组织可以约束经验的分布，却尚未说明现象经验为什么存在。现象属性于是被设定为一个有待解释的事实。

（三）表征主义提供的分析工具

表征主义（representationalism）试图从经验的内容及其表征关系理解意识。一阶表征理论以德雷茨基（Fred Dretske）和泰伊（Michael Tye）为代表，把感觉经验的性质同其表征内容联系起来。[10,11] 高阶思想理论则将一个状态的意识性与主体对该状态的高阶表征相联系；罗森塔尔（David M. Rosenthal）是其中的代表。[12] 克里格尔（Uriah Kriegel）的自我表征理论主张，有意识状态还以特定方式表征自身。[13]

这些理论彼此存在分歧，也并非都支持幻觉主义。不过，表征关系具有一个关键特征：系统能够表征某种性质，即使对象并不按该内容实例化这种性质。水中的直棍看起来弯曲，棍子却未必弯曲。如果主体也通过表征把握自身状态，那么内省内容是否可能失准，就成为可以提出的问题。只是，表征在原则上可错，尚不能证明某个具体的现象判断已经出错。

二、从第一人称证据到幻觉问题

（一）丹尼特对内省权威的质疑

丹尼特（Daniel C. Dennett）的主要论述早于困难问题的正式提出，因此不宜把他简单理解成对查尔默斯的回应。两者在解释起点上形成鲜明对照：查尔默斯从现象经验的存在出发，丹尼特则追问内省究竟能为传统感受质（qualia）的那些特殊性质提供多强的证据。

《意识的解释》以“笛卡尔剧场”（Cartesian Theater）批评这样一种图景：所有心理加工最终汇集成供内部观察者观看的统一呈现。多重草稿模型（Multiple Drafts Model）转而描述并行、分布且持续修订的信息处理过程。[14]《消解感受质》（“Quining Qualia”）又通过内省难以区分经验变化与判断变化的例子，质疑感受质的绝对内在性和内省透明性。[15]

异现象学（heterophenomenology）保留主体对经验的报告、信念和判断，却拒绝预先承诺报告中出现的对象都具有相应的本体论地位。[14,16] 这并不使第一人称材料失去研究价值；相反，它要求我们解释主体为何这样描述自己。主体确信“我有某种特殊的经验性质”是需要尊重的心理事实，但这一确信能否证明性质本身存在，仍是独立的问题。

（二）弗兰克什对解释任务的改写

弗兰克什把这种对第一人称证据的审查发展为系统的幻觉主义立场。[17,18] 幻觉主义所拒绝的是现象实在论（phenomenal realism）关于特殊现象属性的承诺，而非日常生活中所有心理事实。疼痛影响行动，颜色经验存在差异，主体也会形成有关自身状态的复杂判断；这些都可以保留。争论集中于是否还需要承认内省似乎向我们呈现的那类特殊现象性质。国内相关研究也讨论了感受质的本体论地位及认知现象学提出的质疑。[19]

由此产生了一个新的解释对象（explanandum）。传统困难问题询问物理过程为什么伴随现象属性；幻觉问题则要求说明，假如那些属性并不存在，认知系统为何仍把自身表征为具有它们。这里的“幻觉”也不同于日常知觉幻觉。看见一个不存在的人影，已经预设了某种知觉经验；幻觉主义所说的错误，则需要定位在对心理状态的表征、判断或推理之中。

幻觉主义因此不能止步于宣布内省错误。人为何会形成关于“痛感本身”或“红色的主观质感”的概念？为什么这些判断稳定、难以动摇，甚至被主体当作最直接的知识？如果理论只解释“我很疼”这句话如何被说出，而没有解释它所伴随的第一人称确信（如幻肢痛问题），那么它仍未完成自己的任务。[17] 弗兰克什将本体论上的拒绝转化成一组需要认知科学与哲学共同处理的问题。

（三）意识元问题与第一人称显然性

查尔默斯提出的意识元问题（meta-problem of consciousness）同样研究人为什么会认为意识难以解释、不可还原，以及相关判断由哪些认知机制产生。[20] 它在本体论上保持中立：即使我们解释了这些判断的形成，也可以继续主张现象属性存在。幻觉主义则尝试在不预设特殊现象属性的条件下给出解释，并进一步评估第一人称证据的分量。

卡默勒（François Kammerer）把幻觉主义区分为消除性主张和显现主张：传统现象意识不存在，却显得存在。[21] 后一个主张带来更严格的要求。主体通常不把痛感看成某种推测，而把它当作“就在眼前”的事实；理论需要解释的，是现象性为何具有这种内省上的显然性（introspective obviousness）。这也说明，幻觉问题远不只是语言报告的机制问题。

三、幻觉问题的具体解释方案

（一）“显得存在”究竟是什么意思

“意识是一种幻觉”的说法容易造成误解。倘若幻觉主义以某种本身具有现象性的体验来解释现象性为什么“显得存在”，就会重新引入它试图取消的对象。它必须说明这种显现具体由什么构成：可能是对心理状态的表征，也可能是一组判断、倾向或推理关系。同时，它还必须交代主体如何获得有关特殊现象属性的概念，否则即使说明了判断的发生，也无法说明判断的内容。

强幻觉主义（strong illusionism）否认传统意义上的现象意识得到实例化；弱幻觉主义（weak illusionism）则保留某种现象意识，却否认内省赋予它的部分特征。卡默勒讨论的准现象意识（quasi-phenomenal consciousness）指能够承担感觉辨别、情感效价、行为控制等功能，而不具有传统现象属性的心理状态。[21] 因此，即便采取强版本，也可以承认疼痛与愉悦之间有真实差异；需要说明的是这些差异是否还要求特殊的内在质性。

（二）自我模型、属性归因与内省确信

近期的积极解释方案分别处理幻觉问题的不同环节。格拉齐亚诺（Michael S. A. Graziano）的注意图式理论（Attention Schema Theory）认为，系统为预测和调节注意而建立简化的内部模型，省略实际神经加工的大量细节，并可能把自身注意状态表征为一种简单、统一的“觉知”。[22] 它提供了意识自我模型如何形成的机制解释，同样，其并未说明主体为什么将这种模型理解为不可还原的现象性质。

沙巴森（David Shabasson）的推理理论（inference theory）进一步处理现象属性的归因。[23] 主体并不能通过内省直接读取感觉加工的底层机制，却可能把自己获得的有限信息理解为对经验性质的直接把握。在这一条件下，无意识推理能够参与现象性判断的生成。这种方案的优势，是为从普通心理状态到特殊属性判断的过渡提出具体假说；但推理过程是否足以解释第一人称的强烈确信，仍需进一步考察。

卡默勒的元笛卡尔主义（Meta-Cartesianism）则把问题推进到这种确信本身。[21] 我们不仅判断“我有某种经验”，还倾向于认为自己无需借助其他证据就知道它具有什么性质。按照这一解释方向，主体以某种方式将自身状态表征为能直接揭示现象性的状态。幻觉主义需要解释的便不仅是现象判断如何产生，还有它为何显得直接、自明，并拥有特殊的认识论权威。

这三种方案可按其主要任务加以比较：注意图式理论解释有关觉知的自我模型，推理理论解释现象属性归因，元笛卡尔主义解释归因所具有的第一人称显然性。它们目前还并未组合成一个统一的心理机制，也不都以相同方式承诺幻觉主义；不过，它们使原先笼统的“意识幻觉”分解为若干可以分别讨论、检验和修正的解释问题。

（三）作为开放研究框架的幻觉主义

格拉齐亚诺还将幻觉主义视为较宽泛的哲学框架，认为它可以与全球工作空间理论（GWT）、整合信息理论（IIT）等机制研究发生联系，而无须绑定唯一的神经模型。[24] 这一开放性显然具备很好的解释价值：不同理论可以分别研究信息利用、注意、自我表征和内省判断。当然，兼容多种机制并不是幻觉主义正确性的保证。若竞争的可替代理论也能接受同样的机制结果，就还需要说明如何将它们清晰的区分开来。

四、解释力、认识论祛魅与理论限度

（一）从解释机制到证据评价

幻觉主义面临的核心论证距离，是从“现象判断可以得到自然主义解释”走向“相应现象属性不存在”。这两个命题无法等同。视觉科学或许可以解释我们如何获得关于桌子的知觉信念，却无力于取消桌子本身的实在性感受（或感受的实在性），尽管这是一个经典的贝克莱式经验主义问题，但是现代理论所要求的解释力将一种判断的因果来源和其命题的真假区分开来，即需要经过合理性和实在性论证的双重检验。要削弱第一人称证据，理论还须考察产生这些判断的机制是否能够可靠定位其所声称的性质。

卡默勒的无关性论证（Argument from Irrelevance）试图建立这种联系：若某类现象判断的形成无需相应现象属性参与解释，那么这些判断作为属性存在证据的力量便可能受到削弱。[21] 这一思路的意义，在于把注意力由判断的形成过程转向其证据资格。当然，仅仅找到一条无需诉诸现象属性的解释路径，尚不能排除其他解释，也暂且无法直接推出判断与其对象之间不存在可靠联系。

认识论祛魅（epistemic debunking）因此需要比机制说明更严格的条件。它至少要表明，相关判断的发生与其对象是否存在之间缺少适当的认识论联系，或者判断机制容易系统性误导主体。表征主义能够说明内省在原则上可能出错；具体的祛魅论证还须解释，为何这种可能性在现象属性问题上具有实际的证据意义。这里涉及的，是第一人称显然性应当得到多少信任，而非简单判定主体有没有心理活动。

（二）幻觉主义已经解释了什么

即便保留上述论证限制，幻觉主义的积极成果也不应被低估。它要求理论同时说明感知和行为、自我表征、现象性判断及其特殊的内省确信，而不是把主观显现当作无需进一步分析的起点。注意图式理论为意识自我模

型提出认知机制，推理理论将属性归因纳入因果说明，元笛卡尔主义则把第一人称确定感本身变成研究对象。不同方案虽仍有待整合，但已经为意识研究提供了具有解释力的替代路径。

这种解释力还体现在它改变了理论竞争的方式：现象实在论不能只诉诸经验“看起来如此”，还需说明为什么这种显现具有足够的本体论证据力量；幻觉主义也不能只指出内省可能失准，而必须给出能够解释显现及其内容的机制。双方由此可以围绕更明确的问题展开争论，例如哪些现象判断能够被预测，哪些内省特征与可观察过程相关，以及哪种解释承担了更少未经论证的假设。

（三）剩余的本体论问题

仍有两项困难不能省略。首先，现象性显现究竟由哪些普通心理事实构成？如果它只是判断或表征，理论需要说明为何这些活动会使主体觉得自己直接把握了特殊性质；如果它已包含传统意义上的现象性质，幻觉主义就可能陷入循环。其次，即使成功解释这种显现，祛魅所削弱的究竟是朴素感受质理论中的内在性、不可言说性等特征，还是现象实在论的全部形式，仍需逐一论证。

因此，判断幻觉主义的成败，并不取决于它是否已经严格证明现象属性不存在。作为一种本体论主张，它在论证上或许难以证成（这取决于意识问题的特殊困难，无法证成几乎是所有意识命题的最终结局），但作为解释现象判断、内省表征和第一人称显然性的研究纲领，它已经展示了相当强的组织力和继续发展的空间。本体论立场的择取，似乎只是一个是否趁手的研究出发点的问题。

五、结语

当代幻觉主义的变化，可以从解释对象的重新规定来理解。布洛克与查尔默斯使现象意识成为独立问题；丹尼特审查第一人称证据的权威；弗兰克什将“现象意识为什么显得存在”转化为积极的解释任务。近期关于注意、自我模型、推理和内省显然性的研究，进一步把这一任务推进到可以比较的具体方案。

现有理论尚不足以仅凭机制解释完成对现象属性的严格消除，但它们已说明，关于意识的许多强烈直觉可以成为自然主义研究的对象。后续工作的关键，是更清楚地连接判断形成机制与证据评价，并检验这种连接能够支持多大范围的本体论结论。

参考文献：

[1] Nagel, T., 1974. What Is It Like to Be a Bat? The Philosophical Review, 83, pp.435–450.

[2] Block, N., 1995. On a Confusion about a Function of Consciousness. Behavioral and Brain Sciences, 18, pp.227–247.

[3] Weiskrantz, L., 1986. Blindsight: A Case Study and Implications. Oxford: Oxford University Press.

[4] Tsuchiya, N., Wilke, M., Frässle, S. and Lamme, V.A.F., 2015. No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. Trends in Cognitive Sciences, 19, pp.757–770.

[5] Schlicht, T., 2018. A Methodological Dilemma for Investigating Consciousness Empirically. Consciousness and Cognition, 66, pp.91–100.

[6] Phillips, I., 2018. The Methodological Puzzle of Phenomenal Consciousness. Philosophical Transactions of the Royal Society B: Biological Sciences, 373, 20170347.

[7] Chalmers, D.J., 1995a. Facing Up to the Problem of Consciousness. Journal of Consciousness Studies, 2, pp.200–219.

[8] Chalmers, D.J., 1996. The Conscious Mind: In Search of a Fundamental Theory. New York: Oxford University Press.

[9] Chalmers, D.J., 1995b. Absent Qualia, Fading Qualia, Dancing Qualia. In: T. Metzinger, ed., Conscious Experience. Paderborn: Ferdinand Schöningh, pp.309–328.

[10] Dretske, F., 1995. Naturalizing the Mind. Cambridge, MA: MIT Press.

[11] Tye, M., 1995. Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind. Cambridge, MA: MIT Press.

[12] Rosenthal, D.M., 2005. Consciousness and Mind. Oxford: Oxford University Press.

[13] Kriegel, U., 2009. Subjective Consciousness: A Self-Representational Theory. Oxford: Oxford University Press.

[14] Dennett, D.C., 1991. Consciousness Explained. Boston: Little, Brown and Company.

[15] Dennett, D.C., 1988. Quining Qualia. In: A.J. Marcel and E. Bisiach, eds., Consciousness in Contemporary Science. Oxford: Clarendon Press, pp.42–77.

[16] Dennett, D.C., 2003. Who’s on First? Heterophenomenology Explained. Journal of Consciousness Studies, 10, pp.19–30.

[17] Frankish, K., 2016. Illusionism as a Theory of Consciousness. Journal of Consciousness Studies, 23, pp.11–39.

[18] Frankish, K., 2023. What Is Illusionism? Klēsis - Revue Philosophique, 55.

[19] 邢凯伦. 感受质幻觉论的本体论问题与认知现象学纠偏[J]. 科学技术哲学研究, 2025, 42(1): 39–45.

[20] Chalmers, D.J., 2018. The Meta-Problem of Consciousness. Journal of Consciousness Studies, 25, pp.6–61.

[21] Kammerer, F., 2026. House of Mirrors: The Illusion of Phenomenal Consciousness. Oxford: Oxford University Press.

[22] Graziano, M.S.A. and Webb, T.W., 2015. The Attention Schema Theory: A Mechanistic Account of Subjective Awareness. Frontiers in Psychology, 6, 500.

[23] Shabasson, D., 2022. Illusionism about Phenomenal Consciousness: Explaining the Illusion. Review of Philosophy and Psychology, 13, pp.427–453.

[24] Graziano, M.S.A., 2024. Illusionism Big and Small: Some Options for Explaining Consciousness. eNeuro, 11, ENEURO.0210-24.2024.